证券研究报告

事件点评

上海证券 SHANGHAI SECURITIES

# 英伟达发布新一代 GPU 架构，NVLink 连接技术迭代升级

——人工智能行业跟踪报告

## 增持（维持）

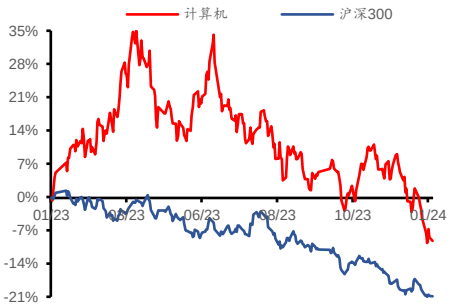
行业： 计算机

日期： 2024年03月29日

分析师： 刘京昭

SAC 编号：S0870523040005

**最近一年行业指数与沪深 300 比较**




## ■ 主要观点

**事件描述**

2024年3月19日，英伟达正式发布Blackwell架构的GPU B200、计算平台HGX B200以及新一代NVLink 5.0连接技术。同时，英伟达基于B200和Grace CPU推出了超级芯片GB200，以及由72张GB200组成的DGX GB200 NVL72超级计算机。

值得关注的是，新一代NVLink连接技术支持单块Blackwell架构的GPU实现1.8TB/s的传输带宽。根据SemiAnalysis的测算，鉴于DGX GB200 NVL72拥有72个OSFP端口，每个端口对应于1个400G或800G光模块，随着GB200数量的增加，网络拓扑结构发生变化，最终GB200对应于800G光模块的数量关系将介于1：2.5到1：3.5之间。

**分析与判断**

我们认为：

（1）以DGX GB200 NVL72为代表的超级计算机，在内部节点间使用铜缆连接，主要是出于降低功耗的考虑，跨机柜连接短期内仍依赖于光收发器。

（2）从生成式AI模型训练需求角度看，跨机柜连接仍为未来主流技术方案，因此数通市场800G光模块需求具备可持续性。

（3）GB200在推理性能上持续优化升级，能够进一步降低生成式AI模型在云侧的推理成本，有助于生成式AI应用在C端落地。

## ■ 投资建议

建议关注：

**中际旭创：**中高端数通市场龙头，2022年与II-VI并列光模块业务营收全球第一。根据iFinD机构一致预期，截至2024年3月25日，公司2024年的预测PE为32倍，位于近五年的93%分位。

**天孚通信：**光器件整体解决方案提供商。根据iFinD机构一致预期，截至2024年3月25日，公司2024年的预测PE为53倍，位于近五年的99%分位。

**新易盛：**光模块领域龙头，成本管控优秀，具备切入增量云计算/AI客户的能力。根据iFinD机构一致预期，截至2024年3月25日，公司2024年的预测PE为42倍，位于近五年的98%分位。

## ■ 风险提示

下游需求不及预期；人工智能技术落地和商业化不及预期；产业政策转变；宏观经济不及预期等。

# 目 录



## 图

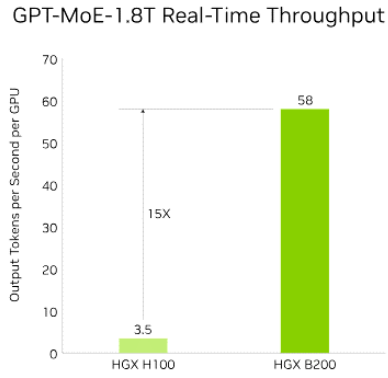


## 表

## 1 英伟达推出 Blackwell 架构，生成式 AI 训练、推理再加速

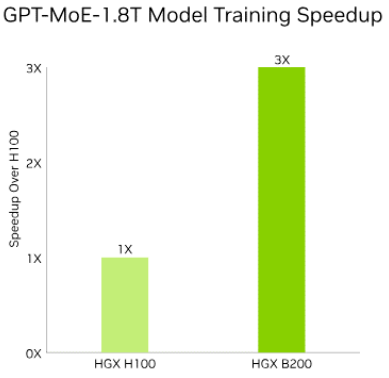
2024 年 3 月 19 日，英伟达正式发布 Blackwell 架构的 GPU B200、计算平台 HGX B200 以及新一代 NVLink 5.0 连接技术。同时，英伟达基于 B200 和 Grace CPU 推出了超级芯片 GB200，以及由 72 张 GB200 组成的 DGX GB200 NVL72 超级计算机。

**图 1：HGX B200 计算平台在生成式 AI 推理场景下实时吞吐量大幅上升**




*资料来源：NVIDIA，上海证券研究所*

**图 2：HGX B200 计算平台在生成式 AI 模型训练场景下训练速率提升明显**




*资料来源：NVIDIA，上海证券研究所*

B200 采用台积电的 4 纳米工艺蚀刻而成，通过 NVLink 5.0 将两个独立制造的裸晶（Die）连接整合，内部共有 2080 亿个晶体管。单个 Blackwell Die 的浮点运算能力相较于 Hopper Die 提高近 25%，总性能提升 2.5 倍，在处理 FP4 精度的浮点运算时，性能还能进一步提升至 H100 的 5 倍。

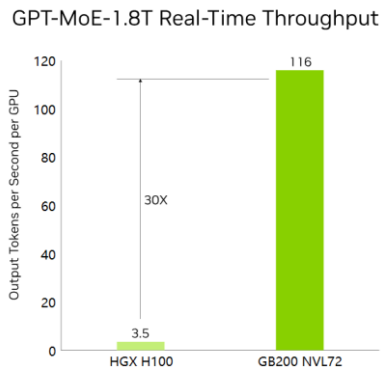
**表 1：Blackwell 架构 GPU 更注重 FP8 和 FP4 浮点运算**

| 技术参数 | HGX B200 | HGX B100 | HGX H200 | HGX H100 |
|---|---|---|---|---|
| GPU 数量 | 8-GPU | 8-GPU | 8-GPU | 8-GPU |
| FP32 吞吐量（FLOPS） | 18P | 14P | 8P | 8P |
| FP16 吞吐量（FLOPS） | 36P | 28P | 16P | 16P |
| FP8 吞吐量（FLOPS） | 72P | 56P | 32P | 32P |
| FP4 吞吐量（FLOPS） | 144P | 112P | -- | -- |
| 显存 | 1.5TB | 1.5TB | 1.1TB | 640GB |
| NVLink 版本 | 第五代 | 第五代 | 第四代 | 第四代 |
| NVSwitch 版本 | 第四代 | 第四代 | 第三代 | 第三代 |
| NVSwitch GPU-to-GPU 带宽 | 1.8TB/s | 1.8TB/s | 900GB/s | 900GB/s |
| 总带宽 | 14.4TB/s | 14.4TB/s | 7.2TB/s | 7.2TB/s |

*资料来源：NVIDIA，上海证券研究所*

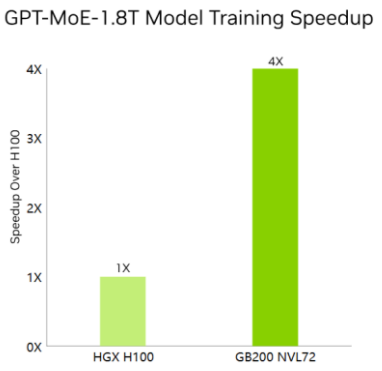
DGX GB200 NVL72 超级计算机包含 18 个 GB200 节点机架和 9 个 NVSwitch 节点机架。每个 GB200 节点搭配 1 个 Grace CPU 和 2 个 GB200 GPU，共计 36 个 Grace CPU 和 72 个 GB200 GPU。在生成式 AI 训练场景下，GB200 NVL72 可支持 720 PFLOPS 的 FP8 吞吐量；在推理场景下，GB200 NVL72 可支持 1.44EFLOPS 的 FP4 吞吐量。

**图 3：GB200 NVL72 在推理场景下实时吞吐量较 HGX100 提升更明显**




*资料来源：NVIDIA，上海证券研究所*

**图 4：GB200 NVL72 在生成式 AI 模型训练场景下性能有所提升**




*资料来源：NVIDIA，上海证券研究所*

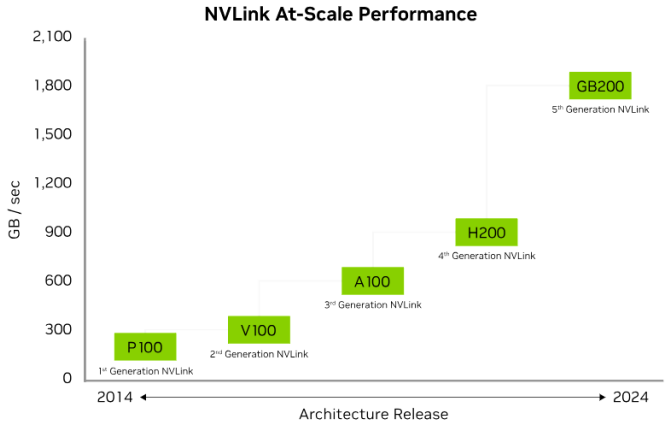
DGX GB200 NVL72 使用水冷散热，在功耗方面，由于使用了 5000 条左右总长度 2 英里的 NVLink 铜缆，在内部的 GB200 节点和 NVSwitch 节点间通信不再依赖光收发器，从而降低近 20KW 的功耗。

**图 5：GB200 由铜缆连接 GB200 节点机架与 NVSwitch 机架**

*资料来源：36 氪，上海证券研究所*

DGX GB200 NVL72 使用第五代 NVLink 实现互联，NVLink 多节点 all-to-all 带宽达到 130TB/s。新一代的 DGX SuperPOD 可由 8 台或 8 台以上的 DGX GB200 超级计算机构成，用户可通过 NVLink 连接 8 台 DGX GB200 超级计算机的 576 块 GB200 GPU，从而进一步扩增集群的共享显存，适应新一代生成式 AI 模型的训练需求。据英伟达介绍，此前需要 8000 块 H100 GPU 使用 90 天时间对 GPT-MoE-1.8T 进行训练，如今只需要 2000 块 GB200 GPU 进行训练，且能耗为使用 H100 训练的四分之一。

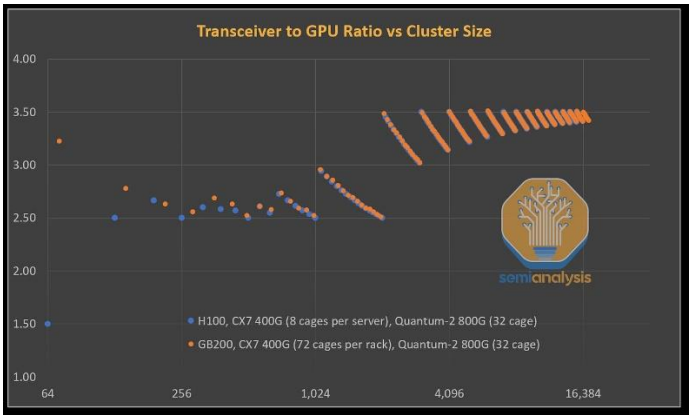
**图 6：GB200 使用第五代 NVLink 连接技术增强 GPU 卡间互联能力**




*资料来源：NVIDIA，上海证券研究所*

值得关注的是，第五代 NVLink 连接技术支持单块 Blackwell 架构的 GPU 实现 1.8TB/s 的双向带宽。根据 SemiAnalysis 的测算，鉴于 DGX GB200 NVL72 拥有 72 个 OSFP 端口，每个端口对应于 1 个 400G 或 800G 光模块，随着 GB200 数量的增加，网络拓扑结构发生变化，最终 GB200 对应于 800G 光模块的数量关系将介于 1：2.5 到 1：3.5 之间。

**图 7：GB200 和 800G 光模块的数量关系与 H100 类似**




*资料来源：SemiAnalysis，上海证券研究所*

我们认为：（1）以 DGX GB200 NVL72 为代表的超级计算机，在内部节点间使用铜缆连接，主要是出于降低功耗的考虑，跨机柜连接短期内仍依赖于光收发器。（2）从生成式 AI 模型训练需求角度看，跨机柜连接仍为未来主流技术方案，因此数通市场 800G 光模块需求具备可持续性。（3）GB200 在推理性能上持续优化升级，能够进一步降低生成式 AI 模型在云侧的推理成本，有助于生成式 AI 应用在 C 端落地。

**表 2：人工智能领域相关公司对比表**

| 所属板块 | 股票代码 | 股票简称 | 22 营业收入 | 22 归母净利润 | 24E 营业收入 | 24E 归母净利润 | 24E 估值 | 近五年 PE 分位数（%） |
|---|---|---|---|---|---|---|---|---|
| 算力 | 688041.SH | 海光信息 | 51.25 | 8.04 | 83.74 | 16.76 | 111 | 77 |
| | 688256.SH | 寒武纪 | 7.29 | -12.57 | 15.38 | -5.98 | -- | -- |
| | 300474.SZ | 景嘉微 | 11.54 | 2.89 | 12.24 | 2.67 | 145 | 92 |
| | 688521.SH | 芯原股份 | 26.79 | 0.74 | 30.96 | 0.29 | 368 | -- |
| | 603019.SH | 中科曙光 | 130.08 | 15.44 | 169.39 | 24.61 | 31 | 49 |
| PCB | 002463.SZ | 沪电股份 | 83.36 | 13.62 | 110.48 | 19.95 | 29 | 93 |
| 光模块/光器件 | 300308.SZ | 中际旭创 | 96.42 | 12.24 | 239.02 | 40.72 | 32 | 93 |
| | 300502.SZ | 新易盛 | 33.11 | 9.04 | 52.86 | 13.07 | 42 | 98 |
| | 300394.SZ | 天孚通信 | 11.96 | 4.03 | 32.72 | 11.63 | 53 | 99 |
| 光芯片 | 688498.SH | 源杰科技 | 2.83 | 1.00 | 2.95 | 1.00 | 130 | 66 |
| 液冷 | 872808.BJ | 曙光数创 | 5.18 | 1.17 | 8.57 | 1.98 | 51 | 89 |
| | 002837.SZ | 英维克 | 29.23 | 2.80 | 54.01 | 5.26 | 34 | 51 |
| 服务器/交换机 | 301165.SZ | 锐捷网络 | 113.26 | 5.50 | 151.68 | 7.89 | 29 | 94 |
| | 603118.SH | 共进股份 | 109.74 | 2.27 | 114.13 | 5.03 | 16 | 96 |
| | 301191.SZ | 菲菱科思 | 23.52 | 1.95 | 33.04 | 2.91 | 23 | 82 |
| | 601138.SH | 工业富联 | 5118.50 | 200.73 | 5442.48 | 252.14 | 19 | 97 |
| | 000938.SZ | 紫光股份 | 740.58 | 21.58 | 904.19 | 27.46 | 22 | 37 |
| | 000628.SZ | 高新发展 | 65.71 | 1.99 | -- | -- | -- | 99 |
| | 600100.SH | 同方股份 | 237.61 | -7.72 | -- | -- | -- | -- |
| | 000034.SZ | 神州数码 | 1158.80 | 10.04 | 1316.69 | 14.60 | 15 | 44 |
| 机器视觉 | 002920.SZ | 德赛西威 | 149.33 | 11.84 | 265.02 | 21.55 | 29 | 6 |
| | 002415.SZ | 海康威视 | 831.66 | 128.37 | 989.19 | 168.70 | 18 | 28 |
| | 002236.SZ | 大华股份 | 305.65 | 23.24 | 374.09 | 43.83 | 14 | 74 |
| | 688003.SH | 天准科技 | 15.89 | 1.52 | 24.38 | 2.74 | 26 | 12 |
| AI+应用 | 300802.SZ | 矩子科技 | 6.84 | 1.29 | -- | -- | -- | 16 |
| | 300418.SZ | 昆仑万维 | 47.36 | 11.53 | 55.89 | 9.64 | 50 | 93 |
| | 688111.SH | 金山办公 | 38.85 | 11.18 | 58.47 | 16.90 | 86 | 29 |
| | 002230.SZ | 科大讯飞 | 188.20 | 5.61 | 256.05 | 13.18 | 87 | 97 |
| | 600570.SH | 恒生电子 | 65.02 | 10.91 | 90.87 | 21.05 | 23 | 2 |
| | 300033.SZ | 同花顺 | 35.59 | 16.91 | 40.43 | 16.90 | 44 | 58 |
| | 600845.SH | 宝信软件 | 131.50 | 21.86 | 195.57 | 33.08 | 28 | 1 |

***资料来源：iFinD，上海证券研究所***

**盈利预测来自 iFinD 机构一致预期；仅列举各板块部分标的；估值基于 2024 年 3 月 25 日收盘价；单位：亿元。*

## 2 风险提示

下游需求不及预期；人工智能技术落地和商业化不及预期；产业政策转变；宏观经济不及预期。

## 分析师声明

作者具有中国证券业协会授予的证券投资咨询资格或相当的专业胜任能力，以勤勉尽责的职业态度，独立、客观地出具本报告，并保证报告采用的信息均来自合规渠道，力求清晰、准确地反映作者的研究观点，结论不受任何第三方的授意或影响。此外，作者薪酬的任何部分不与本报告中的具体推荐意见或观点直接或间接相关。

## 公司业务资格说明

本公司具备证券投资咨询业务资格。

## 投资评级体系与评级定义

| | | |
|---|---|---|
| **股票投资评级：** | 分析师给出下列评级中的其中一项代表其根据公司基本面及（或）估值预期以报告日起 6 个月内公司股价相对于同期市场基准指数表现的看法。 | |
| | 买入 | 股价表现将强于基准指数 20%以上 |
| | 增持 | 股价表现将强于基准指数 5-20% |
| | 中性 | 股价表现将介于基准指数±5%之间 |
| | 减持 | 股价表现将弱于基准指数 5%以上 |
| | 无评级 | 由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级 |
| **行业投资评级：** | 分析师给出下列评级中的其中一项代表其根据行业历史基本面及（或）估值对所研究行业以报告日起 12 个月内的基本面和行业指数相对于同期市场基准指数表现的看法。 | |
| | 增持 | 行业基本面看好，相对表现优于同期基准指数 |
| | 中性 | 行业基本面稳定，相对表现与同期基准指数持平 |
| | 减持 | 行业基本面看淡，相对表现弱于同期基准指数 |
| 相关证券市场基准指数说明：A 股市场以沪深 300 指数为基准；港股市场以恒生指数为基准；美股市场以标普 500 或纳斯达克综合指数为基准。 | | |

投资评级说明：

不同证券研究机构采用不同的评级术语及评级标准，投资者应区分不同机构在相同评级名称下的定义差异。本评级体系采用的是相对评级体系。投资者买卖证券的决定取决于个人的实际情况。投资者应阅读整篇报告，以获取比较完整的观点与信息，投资者不应以分析师的投资评级取代个人的分析与判断。

## 免责声明